

Tickling Proactivity: Exploring the Use of Humor in Proactive Voice Assistants

Nima Zargham
zargham@uni-bremen.de
Digital Media Lab
University of Bremen
Germany

Leon Reicherts
l.reicherts.17@ucl.ac.uk
University College London
United Kingdom

Vino Avanesi
avanesi@uni-bremen.de
Digital Media Lab
University of Bremen
Germany

Yvonne Rogers
y.rogers@ucl.ac.uk
University College London
United Kingdom

Rainer Malaka
malaka@uni-bremen.de
Digital Media Lab
University of Bremen
Germany



Figure 1: An example storyboard used in our online questionnaire presenting a scenario in which the voice assistant uses humor to respond to the user.

ABSTRACT

With rapid advances in artificial intelligence and natural language processing, voice assistants are evolving into advanced digital personal assistants capable of complex tasks. As they become more proficient at understanding people's behaviors, preferences, intentions, and surroundings, opportunities for proactive interactions emerge. However, despite their potential benefits, people still find certain proactive agent interactions inappropriate and invasive, such as correcting or nudging the user. This study investigates humor's potential to enhance the desirability of proactive agent comments, given its stress-relieving and acceptance-promoting characteristics. We investigate how infusing humor into VA statements affects

perceptions of appropriateness and desirability in proactive interventions. We designed storyboards showcasing voice assistants' proactive actions in everyday situations and social contexts. Participants ($N = 50$) assessed these scenarios in an online questionnaire across multiple criteria. Our results reveal that humor's impact on proactive statement desirability is contingent on participants' perceptions of voice assistants and their subjective judgment of the humor.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; *Empirical studies in HCI*; Scenario-based design.

KEYWORDS

Conversational Agents, Voice Assistants, Home Assistants, Proactivity, Humor, Computational Humor

ACM Reference Format:

Nima Zargham, Leon Reicherts, Vino Avanesi, Yvonne Rogers, and Rainer Malaka. 2023. Tickling Proactivity: Exploring the Use of Humor in Proactive Voice Assistants. In *International Conference on Mobile and Ubiquitous Multimedia (MUM '23)*, December 3–6, 2023, Vienna, Austria. ACM, New York, NY, USA, 24 pages. <https://doi.org/10.1145/3626705.3627777>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

MUM '23, December 3–6, 2023, Vienna, Austria

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0921-0/23/12...\$15.00

<https://doi.org/10.1145/3626705.3627777>

1 INTRODUCTION

Voice assistants (VAs) are becoming more advanced and capable of handling complex tasks and conversations. They are commonly used for controlling smart home devices, information gathering, entertainment, online shopping, and time management [60]. With the rise of products such as ChatGPT [13] or smart speakers in homes, conversational agents (CAs) are becoming increasingly important as digital personal assistants. VAs like Apple’s Siri, Google Assistant, Microsoft’s Cortana (now Microsoft Copilot), and Amazon’s Alexa are accessible on various devices such as smartphones, tablets, computers, cars, and smart home devices like Apple HomePod, Google Home, or Amazon Echo. As AI, natural language processing, and sensing technologies advance, researchers predict that these systems will become increasingly proactive [23, 40, 55, 64, 71, 103]. In our previous work [103], we defined proactivity of VAs as “agent-initiated interactions which are triggered by events related to the user(s) and their environment, as opposed to user-initiated inquiries or pre-configured actions, such as reminders, alerts, or routines set by the user.” Previous literature has highlighted the opportunities and benefits that proactive VAs can offer to support, probe, or inspire people [64, 93]. Research has shown that people find proactive VAs highly beneficial, specifically in cases of important reminders, time-saving interventions, or emergency support [103]. Despite the benefits that proactivity can bring, there are also potential challenges, in particular concerning privacy [82], lack of interlocutor authenticity [11], or potential loss of agency [103]. Furthermore, in our previous study [103], we witnessed that proactive interventions for correcting people or nudging them for positive behavior change are often perceived as inappropriate and invasive. In general, CAs often fail to meet consumer expectations [76] and are commonly perceived as machine-like, cold, socially inept, untrustworthy, and incompetent [25, 29, 77].

Humor has been shown to be effective in reducing stress [57] and increasing feelings of well-being [52, 53]. Furthermore, research suggests that humor can make difficult or unpleasant information easier to ‘digest’ [27, 45, 62, 74]. Humor has also shown to be an effective tool in persuasion [50, 95]. Recent research has shown that using humor by CAs enhances service satisfaction [77] and can potentially improve user engagement [78].

Current VAs often use humor to keep people engaged and entertained and compensate for performance limitations [28, 30, 47]. Research on the use of humor in VAs recommends that among the common existing systems, Siri is considered the funniest by people [41, 48]. However, the type of humor used and jokes generated by such systems is often perceived as corny, which can break the illusion of human-likeness, leaving people unhappy, frustrated, and disappointed [47, 75], and damage the emotional connection between humans and the agent. The humor of current VAs is primarily communicated through a number of prescribed jokes, often leading to repetition. One of the most critical elements of humor is timing [58]. Central to its effectiveness are the elements of unpredictability and surprise [6, 88]. The essence of humor lies in its well-timed delivery, aligning appropriately with the situation at hand. This requires agents to possess prior knowledge (e.g., about the user and environment), emotional awareness, situational comprehension, and cultural sensitivity, which often entails proactive

actions [101]. Despite all these challenges, previous research highlights that people wish for more humor in VAs, as evident from the requests for jokes from the agent [10]. In a study conducted by Völkel et al. [90], an elicitation study was undertaken to explore users’ expectations in interactions with an ideal voice assistant. The study revealed that proactivity was an aspect that users wished for voice assistants to exhibit, as well as the use of humor in some cases. Yet, despite the variety of studies on humor and the proactivity of VAs, none have specifically explored the potential of using humor for proactive interventions by VAs.

To build on the previous work about proactive interactions of VAs, in this work, we aim to explore how humor can impact the desirability of proactive VA statements. We further examine the elements that highlight the appropriateness of using humor by identifying in which context and environment such agent interventions are desirable.

We pursue the following research questions:

- RQ1:** Can the use of humor by a VA increase the desirability of its proactive interventions?
- RQ2:** In which situations and context can humor be perceived as more appropriate?

To address our research questions, we employed scenarios presented in our previous study [103] that showcase various proactive actions of a voice assistant in a home setting. We modified the voice assistant’s comments in a three-step process to make them humorous, and presented people with two versions of each scenario, once with the use of humor and once without, and asked them to rate the scenarios regarding *usefulness*, *appropriateness*, *invasiveness*, and how likely they think the user in the scenario will *consider* what the VA says.

Our findings indicate that humor did not consistently improve the desirability of proactive interventions, and where it was not perceived as humorous, it had diminishing effects. However, desirability can be increased depending on participants’ perceptions of VAs and their assessment of whether the VA’s humor was actually humorous.

This research addresses the need for more engaging and desirable interactions with VAs. By exploring the use of humor by a proactive voice assistant in a domestic setting, we contribute to understanding when and in which context humor is perceived as desirable. The findings of this study provide insights for designers and developers to create VAs that effectively incorporate humor, leading to an improved user experience with voice assistants. As highlighted by the literature, perception of humor is highly subjective and depends on one’s socio-cultural background [10, 99]. Nevertheless, there is potential that certain social and environmental aspects of humor can be explored collectively for enhanced utilization, ultimately improving the user experience with voice assistants.

2 RELATED WORK

Previous research has examined proactive services in various applications and technologies such as context-aware reminders or recommendations [79, 86], health and mental well-being [4, 44], or self-tracking to improve productivity [37, 96]. This section provides an overview of related work on proactivity in VAs, humor in

human-computer interaction (HCI), and the role of humor in social interactions.

2.1 Proactive Voice Assistants

Extensive research has been conducted on system-initiated (proactive) interactions within spoken dialogue systems [34, 63, 81]. Although previous research has shown that proactive interactions can open up new opportunities for supporting, probing, or inspiring people [93], current commercial smart speakers remain primarily reactive with users initiating interactions and support only a minimal set of proactive features [64]. Proactive interactions have demonstrated their capacity to be beneficial across various domains, aiding and engaging users. A survey conducted by Schmidt and Braunger [71] involving 1,550 participants indicated that proactivity is a highly valued attribute of voice assistants among users. Additionally, a study by Völkel et al. [89], exploring people's envisioned interactions with an ideal voice assistant, revealed that many participants expressed a preference for proactive voice assistant behavior.

However, one of the biggest challenges with these systems, which is critical to the user experience, is the timing of the interventions [1, 55, 64, 103]. Since speech responses demand immediate attention, they can interfere with people's ongoing activities. This is unlike GUI-based alerts, where users can often delay it until they are ready to take action [63]. Several researchers have looked into opportune moments to proactively interact with people [7, 38, 40, 63, 71–73, 94]. Opportune moments for interaction refer to moments where the disruption of the user's current activity is at a minimum level [85]. Even though it is a fairly easy task for humans to assess another person's current activity before initiating a conversation, designing such behaviors for agents is very challenging [33, 67, 85]. In addition to pinpointing opportune moments for proactive interactions, one crucial aspect is how the agent would deliver them [20, 23, 103]. An adequate delivery could sometimes mitigate the negative effects when the timing might not be perfect. One possible approach for delivering proactive interventions might be the use of humor, which, to the best of our knowledge, is yet to be explored.

One of the major barriers to users' acceptance of VAs is the topic of privacy [16, 51, 101, 103]. A study by Lau et al. [43] revealed that many individuals hesitate to embrace smart speakers due to concerns about privacy and a lack of trust in the companies behind these devices. Adapting proactive services necessitates a higher level of context awareness and access to more personal data, intensifying people's privacy concerns even further [55]. This concern is particularly pertinent in a home environment, where emphasizing the importance of user privacy and security becomes paramount. A study by Tabassum et al. [82] showed that, while users perceived proactive services useful, they were uncomfortable with the always-listening nature of such systems.

Reviewing the literature on proactive interventions of voice assistants reveals that despite some proactive behaviors causing discomfort and being viewed as disruptive and invasive [3], people still recognize many benefits associated with these types of interactions. Previous works suggest taking into account individual user factors, including their current physical and emotional state (e.g.,

stress level, sadness, or fatigue), as well as the surrounding environmental and social context, such as the presence of other people or guests, the closeness of relationships, and the nature and sensitivity of ongoing activities, to foster more favorable interactions [55, 103].

The need for VAs to consider the psycho-social context of their operations to minimize disruptions caused by proactive interventions aligns with the approach required for implementing computational humor. This entails a sensitivity to the social context, which will be elaborated upon in the following sub-section.

2.2 Humor in HCI

Humor plays a crucial role in influencing human behavior and promoting positive social interactions across diverse cultures and societies globally [65]. It is a powerful communication tool, allowing individuals to foster connections and navigate social interactions more effectively [26]. Despite the extensive body of literature exploring humor from various disciplines, such as philosophy, literature, and psychology, there remains a lack of consensus regarding a unified theory of humor [65]. Researchers concur that humor represents a cognitive state of joy, often manifested through facial and vocal expressions like smiles and laughter [47]. A previous study suggests that making creative connections, whether understanding jokes or solving math problems, is an innately pleasurable experience [84]. It is recognized as an inherently ambiguous and context-dependent phenomenon, where its interpretation is contingent upon the specific context in which it occurs [17]. Correspondingly, Martin et al. [53] note that four distinct styles of humor are used in human interaction. Two of these are adaptive (*Affiliative* and *Self-Enhancing humor*), and two are maladaptive (*Aggressive* and *Self-defeating humor*). Further studies have supported the existence and impact of these styles across diverse groups [42, 48].

Within the field of human-computer interaction, humor is recognized as a feature that can enhance engagement, usability, and the personification of technology [47, 58, 75]. Moreover, humor has proven effective in facilitating learning, reducing stress, and fostering intrinsic motivation in various contexts [5, 18, 30, 46, 102]. Using humor in machines aims to imbue them with anthropomorphic qualities, creating a sense of relatability and human-like attributes [47, 101]. By incorporating humor, conversational agents strive to connect with users, evoking perceptions of the agents as more human-like and likable [22, 58]. Consequently, humor becomes a means for CAs to foster attachment [47]. It has also been demonstrated that humor can be used to effectively handle situations where the system is unable to respond to users appropriately [10, 48]. Wei et al. [92] found that users find humorous agents more friendly, intimate, and similar to themselves compared to their non-humorous counterparts. Yet, humor is also often acknowledged as one of the most intricate human qualities to replicate in AI agents [58, 59, 101]. Its multifaceted nature makes crafting even a simple joke complex, necessitating various cognitive abilities, including language skills, theory of mind, symbolism, abstract thinking, and social perception. The challenges in teaching computers to comprehend humor stem from its inherent contextual dependencies, encompassing assumptions, morals, attitudes, and taboos deeply ingrained within humanity's history and cultures [31]. Implementing humor in computers involves three fundamental components:

detection (semantic understanding), generation, and delivery [58]. Even though there have been notable advances in these three areas of computational humor, the development of an agent fully capable of recognizing, generating, and using humor is still not achieved [47, 58, 101]. As such, it has been reported that VA companies often employ professional writers to create comedic responses [35, 47, 56]. This suggests that the current state of technology is still not yet at the level where it can produce sufficiently humorous interactions without the help of humans.

Taking a closer look at each of these components confirms this observation. Regarding humor detection, computational algorithms have been developed to identify humor created by humans. Some studies have focused on simpler forms of humor, such as one-liners [66, 83, 87], while others have explored detecting more intricate expressions like sarcasm, which can be challenging even for humans [36, 61, 98]. Concerning humor generation, HCI researchers argue that AI systems still struggle to consistently produce humorous interactions that meet user expectations [48, 58]. However, it must also be noted that recent advancements in generative AI technologies, such as ChatGPT, have shown promising improvements in this area [14]. And finally, the delivery of humor is arguably still the most challenging aspect of computational humor [58]. To deliver humor effectively, agents need to possess substantial background knowledge about the user, their environment, emotional intelligence, and an awareness of social context and culture.

Even though there are several challenges in integrating humor for agents, the literature argues that people wish for more humor in VAs [10]. However, several studies suggest that humor in VAs depends on the individual and is only appreciated by a subset of users [19, 90, 91]. Research by Völkel et al. [90] suggests that the incorporation of humor by a voice assistant is greatly dependent on individual user preferences. The study observed a disparity in user reactions, with some individuals enjoying humor while others disliking it. Consequently, the authors suggest a cautious approach when integrating humor into voice assistant interactions.

2.3 Humor and Social Interactions

As addressed earlier, beyond its entertainment value, humor plays a crucial role in shaping social dynamics, influencing perceptions, and even challenging societal norms [24, 49, 65]. VAs generally exhibit a socially adaptive style of humor, as demonstrated by Kubert and Korshakova [41]. Their study on humor styles employed by VAs revealed that the prevailing style, across all devices, is *affiliative* humor. This humor style seeks to establish connections and foster bonds between individuals [53]. Furthermore, research by Shin et al. [77] has shown that using affiliative humor by chatbots enhances service satisfaction, as opposed to aggressive humor. This aligns with the idea that affiliative humor is not only suitable in terms of psycho-social sensitivity for incorporating humor into VAs, but it also holds the potential for implementing proactive interventions by fostering a social bond between users and agents. In social interactions, humor appreciation is influenced by the group context within which it occurs [21], including the characteristics of the humor initiator [8, 97]. Previous literature emphasizes that the humor initiator's social status and perceived authority influence how their humor is perceived [68]. For instance, humor delivered

by someone in a position of power might be interpreted differently than if a peer presented the same humor.

In societal relationships, social status and power are highly sought after, motivating individuals to maintain or elevate their position within the hierarchy [2]. Previous research underscores humor's influence on social status [26]. Effective humor can elevate status in new and established relationships, while failed attempts, like inappropriate jokes, can harm it [9]. Romero and Cruthirds [68] suggest that self-enhancing humor can foster positive connections with higher-status individuals, aiding in establishing rapport with superiors or groups like upper management.

In human-agent interaction, research indicates that the more social agency attributed to artificial agents, the greater the reactance displayed by users [69, 70]. Social agency refers here to the perception of the VA as being capable of social behavior resembling human-human interaction [69, 80].

These observations underscore the importance of understanding users' perceived social equality attributed to VAs for comprehending how humor is received from VAs to users. The characteristics of the humor initiator and users' perception of the artificial agent's social attributes play a significant role in understanding humor's impact in such interactions.

From our examination of existing literature, we establish the following research hypotheses for our study:

- H1:** The desirability of a proactive intervention is affected depending on how humorous it is perceived.
- H2:** A correlation exists between how people perceive a VA regarding its social equality and how humorous they find its interventions.

3 METHOD

We conducted an exploratory study consisting of an online survey to examine the impact of humorous proactive interventions by a voice assistant in a domestic setting. Drawing inspiration from scenario-based design methods [15, 64, 103], we used a series of hypothetical storyboards and asked participants to reflect upon and evaluate them. This approach allows us to investigate upcoming technologies despite existing technological constraints. We utilized graphical storyboards to better visualize the situation and spatial configuration of the specific home environment, the user(s), and the smart speaker within the home environment. The efficacy of this method in gaining a good understanding of user perceptions has been demonstrated previously [64, 100, 103].

3.1 Storyboards

We used the scenarios created in our previous study [103] as our initial reference point, depicting proactive VA interventions in a home environment. From their final selection of nine scenarios, we identified five that fell within a moderate range of appropriateness and usefulness and used them as neutral variants for our evaluation.

We employed a three-step approach to design the humorous versions of the scenarios. It is important to note, our aim was not to produce a version of each scenario that would reliably be perceived as humorous by every single individual. It would be unrealistic to try to do so given humor's highly subjective and context-dependent

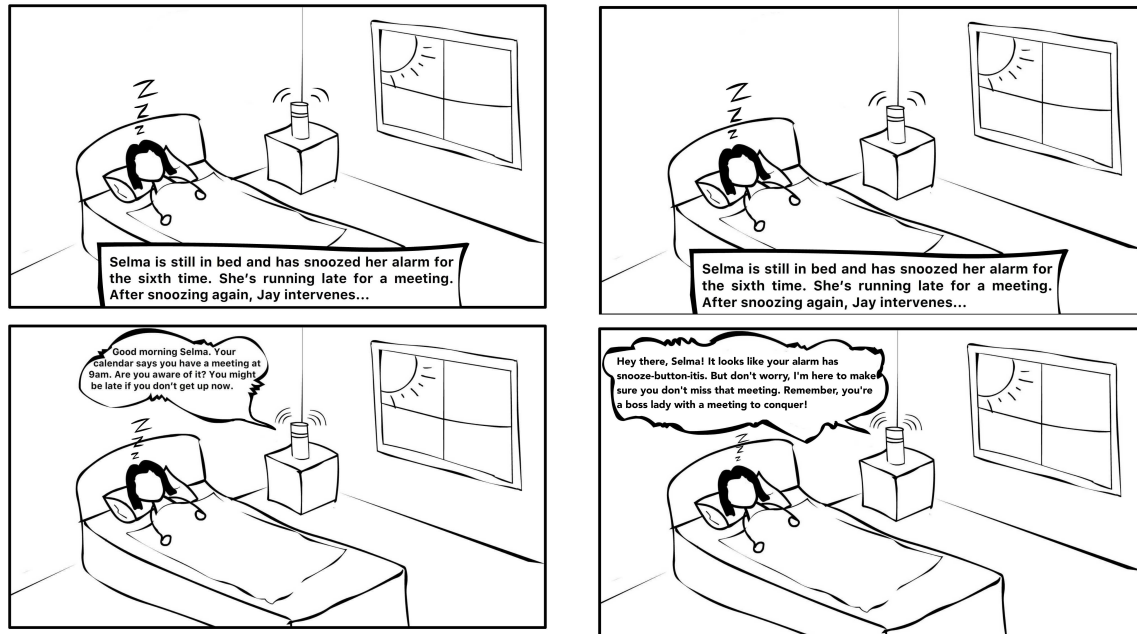


Figure 2: Both versions of the scenario *Meeting Reminder (S1)*. On the left, the neutral version, and on the right, the humorous version is shown. Both versions were evaluated in the survey.

nature. Thus, our aim was rather to produce a version for each scenario that would most likely be seen as *more* humorous than the initial/neutral version of the scenario *on average*. Regarding the type of humor, we exclusively utilized affiliative humor for the agent's comments. As discussed in the related work, prior studies indicate that employing this kind of humor could improve user satisfaction, in contrast to aggressive humor [77]. Initially, for each selected scenario, we generated ten humorous comments using ChatGPT [13]. These generated comments were reviewed by three authors, who assessed their humor and chose their favourite five comments for each scenario. The comments were further refined in an attempt to make them more humorous. Lastly, we presented the selected and refined five humorous comments along with their respective scenarios to a panel of four HCI researchers who were not involved in this project. Based on their feedback, one humorous comment was chosen for each scenario, with some of the selected comments undergoing further modification based on the group's input. This additional step of filtering and refinement by the panel was done with the aim to increase the likelihood that the interventions could be seen as humorous by a wider population in the following study.

Moreover, we included two additional scenarios from the final set of nine scenarios in our previous study [103]. However, unlike the other scenarios, we intentionally left the remark by the agent blank, allowing participants to come up with their own proactive VA interventions.

The scenarios were presented in the form of two-panel cartoon sketches. The design of the storyboards aimed to minimize cultural and ethnic cues to ensure participants could relate to the characters

regardless of their backgrounds. To avoid any potential influence on participants' interpretation of the scenarios, the characters were intentionally designed without facial expressions. Consistent with the original storyboards, the VA in the sketches had a cylinder-shaped appearance resembling a conventional smart speaker. To reduce gender bias, the fictional agent was given the gender-ambiguous name "Jay". For the complete selection of storyboards used in the questionnaire, please refer to the Appendix.

Here is a brief description of each of the scenarios. Both a neutral and a humorous version of each scenario were used in the questionnaire:

- *S1 Meeting Reminder*: After the user has repeatedly "snoozed" the alarm, Jay reminds her of an upcoming meeting.
- *S2 Health Risk*: From the sound of the cough, Jay suspects an elderly user to have a respiratory infection and offers to arrange a doctor's appointment.
- *S3 Fact Checking*: Three friends are discussing a historical topic when Jay interrupts them to get a fact right.
- *S4 Disagreement Clarification*: Two people remember differently about what they agreed on when Jay settles the disagreement by quoting what they said.
- *S5 Nudging*: When the user asks Jay to play a TV series, Jay suggests stopping earlier than last night.

Moreover, here is a brief description of the scenarios where participants had to fill in the agent's proactive comment:

- *S6 Cooking Inspiration*: Two friends are deciding about dinner when Jay proactively intervenes.

- *S7 Technical Support*: A person asks their friend for help setting up new headphones, but the friend is busy cooking. Jay proactively intervenes.

3.2 Online Questionnaire

Participants' responses were collected through the online survey platform Qualtrics¹. The questionnaire began with a welcome text and a brief introduction about the procedure and the research purpose. Participants were then informed about their rights and were required to provide informed consent before proceeding. Afterward, the concept of proactive VAs and the fictional agent "Jay" were introduced to the participants. The initial part of the questionnaire involved participants answering questions about their experience and usage of voice assistants, including their level of interest, enjoyment, and perceived usefulness of VAs. We also provided a clear definition of AI agents and asked participants to indicate their perception of these agents concerning social equality compared to themselves. This evaluation was conducted on a scale of -50 to 50, with 0 representing the agent's equality, -50 representing significant inferiority, and 50 representing significant superiority (to investigate *H2*). Additionally, we asked participants how humorous they would like a VA to be while additionally collecting data on participants' self-assessments of their own humor and how important they find humor in general.

In the subsequent part of the questionnaire, participants were presented with ten scenarios, consisting of five neutral and five humorous scenarios, in a randomized order. Participants were asked to rate each scenario in terms of *usefulness*, *appropriateness*, and *invasiveness*, as well as indicate the likelihood of the user in the storyboard considering the assistant's proposition (following, below: *consideration*). We will in the following generally refer to these variables as the '*four key dimensions*' related to the overall 'desirability' of the interventions (see RQs). Ratings were given using a seven-point Likert scale. Note that higher ratings reflect better perceptions for all four dimensions, including *invasiveness*, for which the scale was inverted to simplify the data analysis and presentation of results (hence, a rating of 1 refers to **most** invasive, and 7 to **least** invasive).

Before this section, participants were informed that, for the purposes of the study, they could assume the fictional agent (Jay) protects their personal data, processes it on the device, and does not share it with any third parties. By pointing this out, we intentionally aimed to alleviate participants' concerns primarily focused on data privacy, as this aspect has been extensively studied in existing research [55, 82]. Participants were encouraged to read the scenarios carefully, as they would be repeated, but the agent's comments would differ.

The next part of the questionnaire involved two scenarios where the agent's comment was left blank. Participants were asked to write their ideal statement for the agent in each scenario and provide their thoughts on the potential impact of a humorous comment from the agent in that context. Participants were then asked to rate the humor of the five humorous scenarios on a Likert scale ranging from 1 to 7 ("How humorous do you find Jay's interaction in this scenario?"). This rating aimed to subsequently examine how the

key dimensions might get affected depending on how humorous participants found the VA's humorous interventions (to investigate *H1*). It was expected that there would be adverse effects on aspects covered by our four key dimensions, such as *appropriateness*, if the humor used should not be perceived as humorous. Subsequently, participants were asked to share their thoughts on the type of humor used in the agent's interventions, including aspects they liked or disliked about the humor. They were also asked to indicate situations where the agent should or should not use humor. The questionnaire concluded with a set of demographic questions about participants' age, gender, nationality, country of residence, and fluency in English.

Prior to running the main study, a pilot study was conducted with two participants. The primary objectives of the pilot study were to identify any potential issues within the questionnaire and assess the scenarios' effectiveness in immersing participants and stimulating contemplation. Subsequently, minor adjustments were made to the questionnaire based on the feedback received, and the main study was conducted. On average, the questionnaire took approximately 20 minutes to complete ($M = 20.03$, $SD = 9.84$). The complete list of questions can be found in the supplementary material.

3.3 Participants

Participants were recruited using convenience sampling, which involved reaching out through mailing lists, social networks, internet forums, and word-of-mouth. Participation in the survey was voluntary and uncompensated. Initially, we obtained a total of 102 responses to our questionnaire. Out of these, 46 responses were excluded due to incompleteness. Furthermore, six participants were excluded from the analysis because their responses consistently lacked informative content, which indicated their unsuitability for our study. These exclusions were made based on their tendency to engage in straight-lining or consistently providing responses that were not pertinent to the questionnaire's content. The final sample consisted of $N = 50$ participants, with 22 identifying as male, 24 as female, three as non-binary, and one participant not specifying their gender. Our study encompassed participants from 16 distinct countries, with the majority residing in the UK (32%), followed by the US (22%), Canada (10%), Germany (8%), Netherlands (8%), and Switzerland (6%). The average age of participants was $M = 33.50$ ($SD = 0.707$). All participants were proficient in English. Among them, 17 have not previously used a voice assistant, while the remaining participants reported rarely (17), sometimes (8), and often (8) using them. 17 participants reported that they own a smart speaker. With regards to participants' self-assessed humor, the items covered *Humor Self* ("How humorous do you think you are?"), *Humor General* ("How important do you find humor in general?"), and *Humor Relationship* ("How important is humor for you in your relationships with other people?"), which participants rated with $Mdn = 5$ ($IQR = 1.25$), $Mdn = 6$ ($IQR = 2$), $Mdn = 6$ ($IQR = 2$), respectively. In addition to the self-assessments, participants rated how humorous they would like a VA to be with $Mdn = 4$ ($IQR = 3$) slightly lower than the previous items.

¹<https://www.qualtrics.com>

Table 1: Medians and IQRs of the sums of participants’ ratings of the four *key dimensions* for the scenarios without humor (baseline) and the scenarios with humor (intervention). On the right side of the table are the *Mann-Whitney U* test statistics; significant results with Bonferroni-corrected $\alpha = .013$ are marked with asterisks. Higher median values are ‘better’ - incl. *invasiveness*, hence higher values mean *less* invasive.

Dimension	Without Humor (Baseline)		With Humor (Intervention)		Wilcoxon Signed-Rank Test Statistics		
	<i>Mdn</i>	<i>IQR</i>	<i>Mdn</i>	<i>IQR</i>	<i>U</i> Statistic	<i>p</i> - value	<i>ES</i>
<i>Usefulness</i>	26.00	7.00	23.00	10.75	795	<.001*	0.680
<i>Invasiveness</i>	15.50	7.75	16.00	9.00	497	0.348	-0.156
<i>Appropriateness</i>	19.00	10.50	17.50	10.00	696	<.001*	0.617
<i>Consideration</i>	22.00	9.75	20.50	9.00	937	<.001*	0.593

3.4 Data Analysis

The questionnaire responses are analyzed and presented both quantitatively and qualitatively to provide a comprehensive understanding of the participants’ views on humorous proactive interventions. These results offer insights into the diverse range of opinions expressed by the participants.

Based on visual inspection of our data and the Shapiro–Wilk statistic, we could not assume normally distributed data. Due to this, as well as the ordinal scale level of most of our items, we conducted non-parametric tests. We used *Spearman* correlations to explore relationships, *Wilcoxon Signed-Rank* tests to compare the difference between baseline and intervention data, and *Mann-Whitney U* tests to compare specific subgroups in our sample. We applied an alpha level of .05 for all our statistical tests.

The open-ended responses were systematically analyzed using a conventional content analysis approach [32]. The analysis began with data familiarisation [12], where two researchers read through all the responses to get a sense of the content and context to understand the patterns, ideas, and concepts present in the responses. Afterward, to develop a coding system, a subset of responses from 10 randomly selected participants were independently coded by two researchers using an inductive coding approach, where a single quote could be assigned to multiple codes, including descriptive (e.g., privacy concerns), conceptual (e.g., benefits of humorous responses), or emotional (e.g., frustration) codes. The researchers engaged in extensive discussions to reach a consensus and establish a coding system. In cases of disagreements, a third author was consulted to ensure agreement. Subsequently, an iterative discussion process between the two authors resulted in the creation of a codebook. One researcher coded the remaining responses individually, employing the established codebook. As the evaluation proceeded, some new codes emerged, requiring the codebook to be adjusted accordingly. This process resulted in extracting key insights and findings from the analyzed responses, presented in section 5.

4 QUANTITATIVE FINDINGS

In this section, we present the quantitative analysis of the questionnaire responses. Variable names are typically presented in *italics*. Descriptive statistics will be reported using *median (Mdn)* and *Interquartile Range (IQR)*. Exceptions are continuous variables like *Age*, for which we will utilize *Mean (M)* and *Standard Deviation (SD)*.

4.1 Perspectives on VAs

To provide a contextual backdrop to our findings, we asked a series of questions from participants regarding their experiences with and attitudes toward VAs. We measured participants’ interest in VAs, their enjoyment while using them, and their perceived usefulness of these systems using a Likert scale ranging from 1 to 7. The participants’ interest in VAs (*Mdn* = 5, *IQR* = 4), enjoyment of using them (*Mdn* = 5, *IQR* = 3), and perceived usefulness (*Mdn* = 4, *IQR* = 2.25) indicate a mostly balanced distribution of general perceptions about VAs among the participants. However, the relatively large IQRs also suggest diverse viewpoints within the sample.

Additionally, we inquired how participants perceived VAs from a ‘social hierarchy’ perspective (“What is your perception of AI agents in comparison to you? – They feel ... to me”). Respondents indicated their perception using a slider with the midpoint representing ‘equality’ (corresponding to a value of 0), the left end signifying ‘highest inferiority’ (corresponding to a value of -50), and the right end representing ‘highest superiority’ (corresponding to a value of 50) in relation to themselves. The *Mdn* = -20 (*IQR* = 32.50) indicates a rating between inferior and equal, slightly ‘leaning towards’ equal. The following sections will refer to this variable as *Social Equality*.

4.2 Comparing the Baseline with the Humorous Scenarios

The scenarios with humor were rated lower than scenarios without humor for *usefulness*, *appropriateness*, and *consideration* – this difference was found to be significant with a *Wilcoxon Signed-Rank* test (see Table 1 for corresponding descriptive and inference statistics). The only dimension that tended to have higher ratings for the scenarios with humor was *invasiveness*; however, the difference was not significant. This suggests that, overall, the humor used by the VA – in the given scenarios – does not seem to affect the four key dimensions positively. The lower ratings of the scenarios with humor could be due to participants not finding the humor used in the scenarios humorous. The humor in the humorous scenarios was rated with *Mdn* = 2.5 for scenario 1 (‘Meeting Reminder’) (*IQR* = 4), and all the other scenarios were rated with *Mdn* = 3 and IQRs ranging from 2.5 to 4 (see Table 3 in Appendix A for descriptive statistics for all scenarios). Overall, this suggests that most participants did not find the scenarios with humor that humorous. However, the high spread (i.e., IQRs) underlines that there are marked individual

Table 2: Medians and IQRs of the *rating deltas* between the scenarios without (baseline) and with humor (intervention) – grouped by participants who found the scenarios more humorous (left side) versus those who found them less humorous (middle). On the right side are the *Mann-Whitney U* test statistics; significant results with Bonferroni-corrected $\alpha = .013$ are marked with asterisks. Higher median values are 'better' - incl. *invasiveness*, hence higher values mean *less* invasive.

Dimension	Above Average Humor Rating		Below Average Humor Rating		Mann-Whitney U Test Statistics		
	<i>Mdn</i>	<i>IQR</i>	<i>Mdn</i>	<i>IQR</i>	<i>W Statistic</i>	<i>p - value</i>	<i>ES</i>
<i>Usefulness</i>	0	5	-6	9	149.0	.002*	0.532
<i>Invasiveness</i>	2	4	-1	5	173.0	.007*	0.446
<i>Appropriateness</i>	0	3	-4	6	240.5	.163	0.230
<i>Consideration</i>	1	4	-6	5	89.5	<.001*	0.714

differences between participants and that they have perceived the humor in the scenarios very differently. This leads to the question of what effects humor might have had on the four key dimensions for participants who found the scenarios **more** humorous compared to those who found them **less** humorous. In other words, in case the VA's intervention is found to be humorous, could this positively affect how *invasive* the intervention is perceived? We will explore this question in the following subsection by investigating how the key dimensions might be affected depending on the participants' humor ratings.

4.3 Effects of Humor When it is Considered Humorous

This section explores how participants' baseline and humorous scenario ratings (for the four key dimensions) differ depending on how humorous they found the latter. To explore this, the sample was split into two halves (ex post) based on their overall ratings (using the sum of humor ratings of all five scenarios for each participant). One group was defined containing all participants above the median (*Mdn* = 16.5) of the humor rating sums (which we will refer to as *Higher Humor Rating Group*, $n = 25$) and the other group below the median (*Lower Humor Rating Group*, $n = 25$). Using the median instead of the scale's midpoint ensured that both sub-samples were equally sized. However, it is important to stress that the upper half does not exclusively comprise participants who found the scenarios humorous overall. This is due to the median ratings being positioned below the midpoint of the 7-point Likert scale (see also Table 3 in the Appendix A).

When inspecting Table 2, it can be seen that the *Lower Humor Rating Group* (who found the scenarios with humor less humorous) consistently rated them worse across all *four key dimensions* than the scenarios without humor – thus presenting a similar picture as in the previous section (subsection 4.2) but with the negative effects being even more pronounced. However, a different picture emerges when considering the *Higher Humor Rating Group*, where there seemed to be no adverse effects on the *four key dimensions* (with rating deltas ranging between 0 to 2) and for *invasiveness* and *consideration* there even seemed to be positive effects (see also Figure 3).

Taken together, the deltas thus were all < 0 for the participants who found the scenarios **less** humorous and ≥ 0 for participants who found the scenarios **more** humorous. To investigate if the

differences between the two groups are significant, a *Mann-Whitney U* test was conducted for each of the four key dimensions, which was significant for *usefulness* ($p = .002$), *invasiveness* ($p = .007$), and *consideration* ($p = < .001$), but not significant for *appropriateness* ($p = .163$).

Given this significant difference in the *invasiveness* and *consideration* rating deltas and since the deltas were positive for the *Higher Humor Rating Group*, an exploratory analysis was conducted to examine if the increase from baseline to intervention is significant for these two dimensions when only considering this group. For *invasiveness* the difference is indeed significant (Wilcoxon Signed-Rank, $W = 47$, $p = .002$, *Effectsize* = -0.687) while for *consideration* it is not ($W = 144$, $p = .579$, *Effectsize* = -0.044).

4.4 Perceived Social Equality of VA

We expected that participants would find the scenarios with humor more humorous if they perceive the VA more *socially equal* to them. Indeed, there seems to be a significant correlation ($p = .043$, $r_{Spearman} = 0.288$). This is further corroborated when examining the *VA Social Equality* ratings of participants who stated that they preferred the scenarios with humor over those without humor ("In general, did you prefer the humorous interactions over the non-humorous ones?"). A marked difference can be observed in participants' *VA Social Equality* ratings for those who prefer the scenarios with humor *Mdn* = 47.5 (*IQR* = 32.5) compared to those who prefer those without *Mdn* = 25 (*IQR* = 21.3), see also Figure 4. This difference was found to be significant using a *Mann-Whitney U* test with $U = 146$, $p = .004$, and *Effect size* = 0.493. This suggests that the more people see VAs at a similar social level to themselves, the more they are open to the VA using humor.

5 QUALITATIVE FINDINGS

Within our sample of 50 participants, nearly half of them (23) expressed their dislike for the style of humor used in the scenarios. They perceived the humor as inappropriate, forced, lacking personal connection, and bothersome. For instance, one participant remarked: "None of the characters in the scenarios were joking around with their friends. I would find the comments irritating if an actual human had made them. Not only is it irritating, but it makes it far less clear what the AI is actually saying or offering to do." (P14). On the other hand, 14 participants embraced the humor,

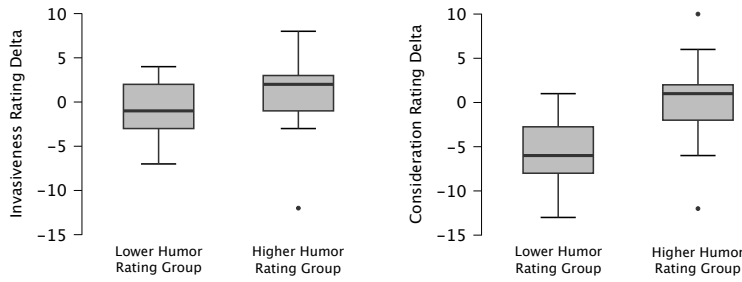


Figure 3: Boxplots of baseline to intervention rating deltas for *invasiveness* and *consideration* grouped by participants below and above the average scenario humor rating, showing median, IQR, and maximum and minimum values (with three outliers represented as dots).

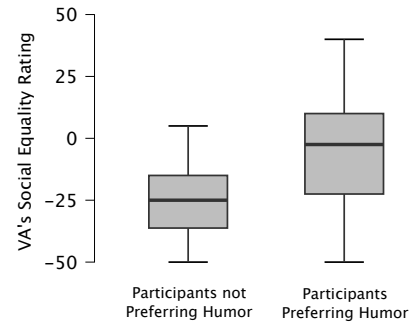


Figure 4: Boxplot of *VA Social Equality* ratings grouped by participants who preferred scenarios with/without humor, showing median, IQR, and max/min.

finding it both enjoyable and intriguing. One participant articulated: “[The humor] makes the intervention more natural.” (P36). Seven participants underscored the subjective essence of humor, acknowledging the challenge of crafting humorous comments for voice assistants.

Moreover, four participants expressed concerns regarding continuous monitoring of the auditory environment by the agent. One participant stated: “It raises security concerns about the constant surveillance of household audio.” (P43). These concerns were raised even though participants were explicitly requested to temporarily set aside privacy and data protection considerations during the survey.

5.1 Humor Ranking

Participants indicated their favorite humorous scenario and provided the rationale behind their choice (see Figure 5).

The *Meeting Reminder* emerged as the favorite among 12 (24%) of participants. Participants found humor in this scenario to be encouraging, a blend of entertainment and utility, as well as inspiring and motivational. One participant mentioned: “Calling the user a ‘boss’ is a colloquial and personable interaction that does not feel forced and is motivating. It is how a friend would speak to you.” (P43).

As for the *Fact Checking* and *Nudging* scenarios, 11 (22%) participants favored them. In the *Fact Checking* scenario, the humorous agent intervention was perceived as ‘funny yet factual’, with enjoyment derived from a historical reference, and an opinion that it alleviates tension. One participant mentioned: “It is humorously delivering the fact while not making the situation unnecessarily awkward.” (P19). In the *Nudging* scenario, participants viewed the comment as both humorous and effective, suitable as it aligned with the entertainment context (two), and not detracting attention from the issue (one): “It brings humor without undermining the seriousness of the matter.” (P5).

Seven participants favored *Disagreement Clarification*, mainly citing its tension-relieving aspect (three). One participant pointed out:

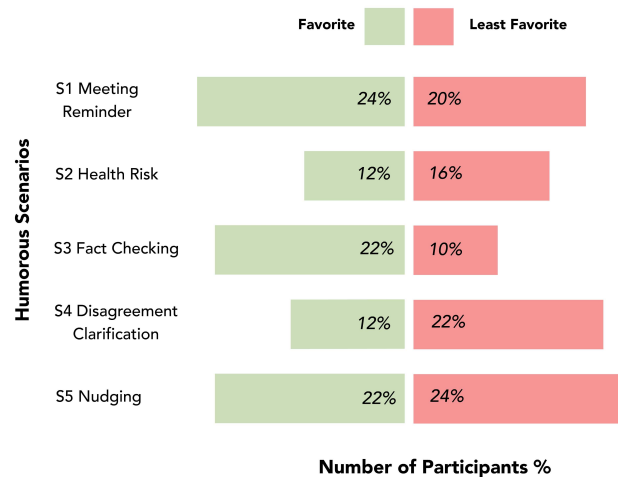


Figure 5: Bar charts displaying the percentage of participants selecting a humorous scenario as their favorite and least favorite for each scenario.

“It breaks up an awkward/tense scenario, and the subject matter is pretty light and inconsequential.” (P33). The *Health Risk* scenario was favored by six participants, primarily as it was perceived as unintrusive: “It is a touch of humor without sounding condescending or juvenile.” (P30). Three participants did not select a favorite scenario.

In terms of participants’ least favorite humorous scenario and the reasoning behind it, the *Nudging* scenario garnered the highest number of votes for being the least favored by 12 people. Participants expressed concerns about the agent’s remark in this scenario being impolite, inappropriate, and overly lengthy. One participant pointed out: “It feels a bit weird that a piece of technology would be questioning what I’m doing.” (P11).

The *Disagreement Clarification* scenario was chosen as least favorite for 11 respondents, due to perceptions of the interaction as intrusive, impolite, and pedantic. A participant highlighted: “[The agent] is negative towards one person.” (P42). Another participant noted that “the agent’s involvement felt intrusive in a personal relationship.”

Meeting Reminder was selected as the least favorite by ten participants. Two found the agent’s behavior insensitive to context, two others thought it is exerting pressure on the user to be productive, and three participants specifically disliked the employed humor. One participant commented: “It comes across as a company trying far too hard.” (P25).

Eight chose *Health Risk* as their least favorite. Participants thought that the humor employed could detract from the gravity of the health concern. Five participants highlighted that humor is inappropriate when dealing with health matters. One participant stated: “It’s not good to add that level of humor into matters related to people’s health, especially when coming from an AI.” (P33).

Finally, *Fact Checking* was chosen by five respondents as the least favorite, primarily due to the comment being perceived as intrusive: “Nobody asked Jay’s opinion. Maybe it could light up to show it has something to contribute.” (P10). Two participants disliked the humor employed in this scenario. Four participants did not select a favorite scenario.

5.2 What Would the Agent Say?

Participants filled in their ideal agent’s comment for the two scenarios of *Cooking Inspiration* and *Technical Support*.

Regarding *Cooking Inspiration*, a significant majority (40 out of 50) offered supportive comments without incorporating humor, by suggesting food ideas or facilitating online food ordering. Three participants expressed the belief that the agent should not engage in such a context. Only five participants chose to introduce humor into their agent’s comment. They either used a humorous food recommendation, or a humorous comment followed by supportive guidance: “Not this again, I can’t remember the last time you knew what to eat. Luckily, I can help you – how about some spaghetti Bolognese?” (P28). Two left this question unanswered. Regarding the impact of humor, within this scenario, 16 people thought that incorporating humor would not have any influence on the situation. In contrast, 15 participants believed that humor might exacerbate the situation, associating it with potential annoyances, distractions, time wastage, diminished seriousness, or elements of irritation and condescension. One participant mentioned: “It would sound more invasive and less like a service.” (P16). On the contrary, eight participants thought that humor could enhance the situation, being seen as ‘encouraging’ or ‘inspiring’ for users. One participant said: “I think humor would make it more light-hearted and pleasant.” (P8). 11 participants did not offer a response to this aspect.

Similarly, regarding *Technical Support*, the majority (32) offered a supportive comment without incorporating humor, three believed the agent should not engage, and five left this question unanswered. For this scenario, ten participants used humor in their comment. Such comments were either a humorous statement, or a humorous statement followed by supportive guidance: “Sandra, let me be the chivalrous one here and help with the headphones.” (P13). About

the impact of humor on the situation for this scenario, 18 found the use of humor to be beneficial, as it could lighten the mood, help release tension, and make the interaction more appropriate. One participant stated: “It diffuses a potentially tense situation by lightening the mood” (P8). 16 people thought it will make the situation worse as it can be annoying, distracting, inappropriate, or it can decrease the seriousness of the situation. One participant said: “It would be inappropriate to joke as everyone is busy.” (P3) Eight participants believed humor would not have any impact on the situation, and ten did not respond to this question.

5.3 Opportune Time for Humor

When considering appropriate times for an agent to employ humor, 15 participants indicated that it should be employed during non-serious and playful instances, such as when people are in a playful mood or laughter is detected. Six people mentioned that humor should be utilized exclusively when explicitly requested by the user. Additionally, three suggested its usage when users are in the company of close friends or family members. Three proposed its application during moments of perceived tension to alleviate stress. Two participants recommended a consistent humorous approach, while one participant suggested leveraging humor to motivate people toward healthier behaviors.

On the contrary, four people expressed a preference for the agent to refrain from using humor altogether. Regarding contexts where participants felt humor should be avoided, half of the people (25) noted that the agent should abstain from using humor during discussions of serious topics such as health, work, or finances. An additional five participants emphasized that humor should not be used during time-critical situations, while another five highlighted that humor should be avoided in socially tense situations.

6 DISCUSSION

Our exploratory investigation delved into incorporating humor in proactive VA statements within a home environment.

We interpreted the results of this evaluation to provide answers to the following comprehensive questions:

RQ1: Can humor increase the desirability of proactive interventions of VAs?

RQ2: In which situations and context is humor perceived as appropriate?

6.1 Impact of Humor Reception on Desirability

Overall, regarding **RQ1**, our questionnaire results demonstrated that the humor used in our scenarios did not affect aspects of *usefulness*, *appropriateness*, *invasiveness*, and *consideration* positively.

However, there were marked differences between the participants. Some found the humor heart-warming and pleasant, while others considered it distracting and inappropriate. This once again underscores the inherent subjectivity of humor [10, 99, 101]. However, we observed that certain factors impacted the desirability of humorous proactive comments by VAs, which we discuss here.

We witnessed that around half of our participants did not like the humor used in our scenarios. For this subset of participants, the inclusion of humor predominantly resulted in a negative influence on the proactive interventions made by the VA in terms of *usefulness*,

appropriateness, *invasiveness*, and *consideration*. On the contrary, another subgroup of participants generally enjoyed the humor incorporated into the scenarios. For this category, humor within our scenarios generally positively impacted the VA's proactive interventions concerning *invasiveness*. Taken together, these findings suggest that when the humor used fails to resonate with users, it is likely to adversely affect the user's perception of the VA's proactive statement. This is in line with prior research indicating that humor carries inherent risks, and if a humorous attempt falls short, it can lead to worse outcomes [8, 9, 39]. Conversely, if the humor used is indeed perceived as humorous by the user, it has the potential to mitigate the *invasiveness* of the comment. In such cases, humor can act as a buffer, making people more receptive to proactive interventions. This aligns with existing literature on humor, suggesting that humor can enhance the reception of information [27, 45, 62, 74]. In effect, we can accept our first hypothesis:

H1: The desirability of a proactive intervention is affected depending on how humorous it is perceived.

6.2 Perceived Social Equality and Humorous Interventions

We witnessed that the effectiveness of delivering a humorous intervention can be heightened when the user perceives the VA as a more socially equal partner. Our evaluation highlighted that participants who viewed the VA as more socially equal tended to rate humorous scenarios as funnier than those who perceived the VA as inferior. Moreover, we observed that the more participants saw VAs at a similar social level to themselves, the more they were open to the use of humor by VAs. These findings are consistent with existing literature, emphasizing that the perceived characteristics of the individual delivering humor impact its reception [8, 97], particularly evident concerning the social status and perceived authority of the individual delivering humor [68]. In line with these insights, we can then confirm our second hypothesis:

H2: A correlation exists between how people perceive a VA regarding its social equality and how humorous they find its interventions.

The implications of these findings suggest that VAs should tailor their use of humor based on the user's perception of their relationship with the VA. This perception could be gathered through user self-reports during VA setup or configuration. Additionally, VAs could adjust their application of humor based on the given context. This involves determining whether the VA should function predominantly as an assistant for task-oriented assistance or as a 'colleague' aimed at motivating and inspiring the user. These distinct roles could imply different hierarchies and user expectations concerning the 'social' interaction and its perceived 'hierarchy'.

6.3 Timing Humorous VA Statements

Regarding **RQ2**, our qualitative assessment showed that participants expressed the belief that VAs should refrain from using humor during discussions or activities related to serious topics such as health, work, or finances. Additionally, participants emphasized the importance of avoiding humor in time-critical and socially tense situations. These findings underscore the significance of timing and

context in deploying humor. The least favored scenarios further shed light on this matter. Participants expressed disapproval when humor was not carefully contextualized, leading to perceptions of impoliteness and inappropriateness. Moreover, people raised concerns when using humor in contexts involving sensitive or serious topics, worrying that it might undermine the gravity of the subject matter. Although humorous content can be attention-grabbing and entertaining, it might also convey that a situation is not serious [54]. Humor could potentially lead to a reduced inclination to address a problem due to its association with non-serious contexts. This was also the case in our findings, where *usefulness* was generally rated lower in the scenarios with humor even though the type of help or suggestion was not different and thus the 'objective usefulness' technically being the same.

Participants preferred humor during light-hearted and playful occasions. They suggested that humor could be appropriately employed when cues like laughter or humorous conversations are detected, signaling an opportune moment for the VA to engage in humor. Another factor was regarding the people's relationship, proposing using humor when people are with close friends or family members. The favored scenarios shed further light on this aspect. Participants preferred humorous VA comments that strike a balance between entertainment and utility, fostering a motivating and encouraging atmosphere. Additionally, we observed a potential for using humor to alleviate tension and enhance user experience, particularly when combined with factual information and contextual relevance, to create relatable and positive interaction dynamics.

Participants generally disliked the use of humor in the *Disagreement Clarification* scenario due to its perceived tense social context and in the *Health Risk* scenario due to the potential seriousness of the health concern. In contrast, regarding *Fact Checking*, participants seemed to find the context suitable and the topic less serious, resulting in a more favorable reception of humor. In the case of *Meeting Reminder* and *Nudging*, opinions were rather mixed regarding the appropriateness of humor.

Nevertheless, some participants favored a reserved approach, desiring the agent to deploy humor only upon specific requests. Conversely, a group endorsed a consistent use of humor, valuing a consistent presence of humor in interactions. As for the potential impact of humor concerning the *fill-in-the-blank* scenarios, participants displayed a range of opinions. Some found it beneficial, some perceived it as having no influence, and others believed it could worsen the situation. This variation underscores the subjective nature of humor's effects and its nuanced reception across different individuals and contexts. Participants' diverse viewpoints highlight the intricate nature of deploying appropriate humor. The findings underscore the complexity of humor and the necessity of factoring in user preferences, context, and potential impacts when incorporating humor into VA interactions, as mentioned in previous research [101].

The significant preference for supportive comments without humor regarding both *fill-in-the-blank* scenarios further suggests that participants value straightforward and pragmatic interactions. Even when humor was used, it often accompanied supportive guidance, revealing a desire for practical assistance alongside any attempt at humor.

An important observation from the qualitative evaluation of humorous scenarios was that participants directed their attention mainly toward the proactive intervention itself and its timing, overseeing the humorous aspect of the agent's comment. This highlights that the novel interaction introduced by the agents' proactive statements took precedence, often overshadowing the humor intended. Such a pattern of responses could imply that when participants favored a humorous approach, the success could be attributed to the fitting and appropriate timing of the proactive intervention 'itself'. This observation suggests that the timing of proactive interventions may align with suitable moments for incorporating humor.

6.4 Humor and Proactive VA Desirability

Our findings highlight the intricacies of integrating humor into proactive voice assistant interactions. If humor fails to resonate with the user, it can have counterproductive consequences, especially in the context of proactive VA interventions. It became evident that humor is not a mere supplementary aspect or interactional feature that can be casually incorporated. However, it could enhance the interaction if it resonates with the user. To this end, we recommend tailoring humor to individual user preferences and sensitivities. This approach acknowledges the diverse reactions that humor can elicit among people. For designers and developers of VAs, understanding that humor can have varying effects on users is crucial. Therefore, investing in implementing personalized humor that resonates with users' unique perspectives is a worthwhile consideration. Overall, designers should consider humor as a potential strategy to soften the impact of proactive interventions. However, if humor cannot be achieved and tailored to individuals, alternative approaches might be more effective in achieving desirable outcomes.

7 LIMITATIONS AND FUTURE WORK

Our research has certain limitations that require acknowledgment. Firstly, even though our study had a heterogeneous sample with varying ages and backgrounds, the findings should be interpreted within the specific group studied. Our 50 participants resided in 16 different countries. While our sample included individuals with various cultural backgrounds, it is important to note that the sample size remains relatively small and might not offer a fully representative picture. Prior literature has underscored the influence of cultural background on humor interpretation [10, 99]. To enhance the robustness of our findings, future research should extend its investigation to broader and more varied populations. Furthermore, as we addressed earlier, humor perception is inherently subjective [10, 101]. Enhancing the desirability of humorous agents' comments requires a deeper comprehension of users' individual preferences, personalities, and cultural influences, as well as one's individual 'history' with an agent. Subsequent studies could delve into crafting personalized humorous remarks aligned with each user's humor taste. In this work, we employed a three-step approach in an attempt to produce scenarios that would, on average, be perceived as more humorous than the baseline. Our results indicated that a significant portion of our participants did indeed perceive the scenarios as humorous. However, another subset of our participants did not share the same perception about the humor level in the humorous scenarios. It is essential to acknowledge that, due to the

inherent subjectivity of humor, it is not possible to ensure that all participants will find all the scenarios humorous. Nevertheless, this was not a major issue for our study design, which accounted for some differences in humor perceptions.

Our study explored humorous proactive VA comments within a home environment, as it is one of the most common use cases for VAs. While the broader insights from this study may have applicability in other settings, future research should delve into these VA remarks within different contexts, such as workplaces and public spaces. Moreover, based on recommendations from previous literature, we only employed affiliative humor for the agent's humorous comments, as this form of humor has been shown to enhance service satisfaction as opposed to aggressive humor [77]. In future studies, other types of humor should also be examined to understand their impact on user experience. Humor is a phenomenon greatly influenced by context and timing. In our approach, we made an effort to incorporate context and timing within our storyboards to a certain extent. However, storyboards cannot convey the exact turn-taking, timing, and delivery in a given (social) context that might play a role in how (humorous) an intervention is perceived. Thus, future studies exploring humor for VAs may consider alternative methods that can more effectively capture and utilize these crucial elements.

Lastly, it is important to acknowledge the limitations of our chosen method in this study. We gathered people's opinions based on hypothetical scenarios, as many of the capabilities depicted in our storyboards are not currently present in consumer products. This approach enabled participants to engage in speculation about interactions with future technologies that might be complex or costly to develop. Nonetheless, we must acknowledge that participants did not directly experience these situations, and their perceptions might not fully align with real-world experiences.

8 CONCLUSION

In this study, we explored the utilization of humor in proactive voice assistants and its influence on the desirability of such interactions. We conducted an online questionnaire with 50 participants, employing a scenario-based method. Participants were presented with storyboards illustrating instances where a proactive smart speaker engaged with people in various everyday situations, utilizing humorous and non-humorous remarks. Our results reveal that, while humor did not uniformly enhance aspects of *usefulness*, *appropriateness*, *invasiveness*, and *consideration*, there were clear distinctions in participants' reactions, highlighting the subjectivity of humor. We witnessed that humor's effects on the desirability of an intervention depend on whether people perceive it as humorous or not. Additionally, the success of humorous interventions can be enhanced when people perceive the VA as more socially equal. We recommend personalized humor tailored to individual user preferences and sensitivities to address these diverse responses. Humor is a multifaceted tool, with its effects contingent on individual preferences, context, and perceptions. Our findings caution against the casual incorporation of humor. Instead, humor should be applied thoughtfully or avoided altogether, as misaligned humor can backfire, particularly within the context of proactive VA interventions. Recognizing the significant role humor has historically played in

human social interactions and relationships from the origins of society, we contend that a proper understanding and exploration of humor in the context of human-computer interaction should be encouraged in both research and practical endeavors within this domain [58, 101].

ACKNOWLEDGMENTS

We would like to sincerely thank Paweł W. Woźniak for supporting us with parts of the data analysis. We also thank the anonymous reviewers whose suggestions helped improve and clarify this work. This work was partially funded by the Klaus Tschira Foundation, by the Leverhulme Trust (award DS-2017-026), by the FET-Open Project 951846 “MUHAI – Meaning and Understanding for Human-centric AI” funded by the EU program Horizon 2020, as well as the German Research Foundation DFG as part of Collaborative Research Center (Sonderforschungsbereich) 1320 “EASE – Everyday Activity Science and Engineering”, University of Bremen (<http://www.ease-crc.org/>).

REFERENCES

- [1] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3290605.3300233>
- [2] Cameron Anderson, John Angus D Hildreth, and Laura Howland. 2015. Is the desire for status a fundamental human motive? A review of the empirical literature. *Psychological bulletin* 141, 3 (2015), 574.
- [3] Kika Arias, Sooyeon Jeong, Hae Won Park, and Cynthia Breazeal. 2020. Toward Designing User-centered Idle Behaviors for Social Robots in the Home.
- [4] Daisuke Asai, Jarrod Orszulak, Richard Myrick, Chaiwoon Lee, Joseph F Coughlin, and Olivier L De Weck. 2011. Context-aware reminder system to support medication compliance. In *2011 IEEE international conference on systems, man, and cybernetics*. IEEE, New York, USA, 3213–3218.
- [5] Oswald Barral, Ilkka Kosunen, and Giulio Jacucci. 2017. No Need to Laugh Out Loud: Predicting Humor Appraisal of Comic Strips Based on Physiological Signals in a Realistic Environment. *ACM Trans. Comput.-Hum. Interact.* 24, 6, Article 40 (Dec. 2017), 29 pages. <https://doi.org/10.1145/3157730>
- [6] Nancy Bell and Anne Pomerantz. 2014. Reconsidering language teaching through a focus on humor. *EuroAmerican Journal of Applied Linguistics and Languages* 1, 1 (2014), 31–47.
- [7] André Berton, Dirk Bühler, and Wolfgang Minker. 2006. SmartKom-Mobile Car: User Interaction with Mobile Services in a Car Environment. In *SmartKom: Foundations of Multimodal Dialogue Systems*, Wolfgang Wahlster (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 523–537. https://doi.org/10.1007/3-540-36678-4_33
- [8] T Bradford Bitterly. 2022. Humor and power. *Current Opinion in Psychology* 43 (2022), 125–128.
- [9] T Bradford Bitterly, Alison Wood Brooks, and Maurice E Schweitzer. 2017. Risky business: When humor increases and decreases status. *Journal of personality and social psychology* 112, 3 (2017), 431.
- [10] Pavel Braslavski, Vladislav Blinov, Valeria Bolotova, and Katya Pertsova. 2018. How to Evaluate Humorous Response Generation, Seriously?. In *Proceedings of the 2018 Conference on Human Information Interaction and Retrieval* (New Brunswick, NJ, USA) (CHIIR '18). Association for Computing Machinery, New York, NY, USA, 225–228. <https://doi.org/10.1145/3176349.3176879>
- [11] Michael Braun, Anja Mainz, Ronee Chadowitz, Bastian Pflöging, and Florian Alt. 2019. At your service: Designing voice assistant personalities to improve automotive user interfaces. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–11.
- [12] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.
- [13] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [14] Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4. arXiv:2303.12712 [cs.CL]
- [15] John M. Carroll. 1999. Five Reasons for Scenario-Based Design. In *Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences - Volume 3 - Volume 3 (HICSS '99)*. IEEE Computer Society, USA, 3051.
- [16] Narae Cha, Auk Kim, Cheul Young Park, Soowon Kang, Mingyu Park, Jae-Gil Lee, Sangsu Lee, and Uichin Lee. 2020. Hello There! Is Now a Good Time to Talk? Opportune Moments for Proactive Interactions with Smart Speakers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 74 (Sept. 2020), 28 pages. <https://doi.org/10.1145/3411810>
- [17] William Curran, Gary J McKeown, Magdalena Rychlowska, Elisabeth André, Johannes Wagner, and Florian Lingenfels. 2017. Social Context Disambiguates the Interpretation of Laughter. *Frontiers in psychology* 8 (2017), 2342. <https://doi.org/10.3389/fpsyg.2017.02342>
- [18] Claire Dormann and Robert Biddle. 2006. Humour in game-based learning. *Learning, Media and Technology* 31, 4 (2006), 411–424.
- [19] Philip R. Doyle, Justin Edwards, Odile Dumbleton, Leigh Clark, and Benjamin R. Cowan. 2019. Mapping Perceptions of Humanness in Intelligent Personal Assistant Interaction. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services* (Taipei, Taiwan) (MobileHCI '19). Association for Computing Machinery, New York, NY, USA, Article 5, 12 pages. <https://doi.org/10.1145/3338286.3340116>
- [20] Mateusz Dubiel, Kerstin Bongard-Blanchy, Luis A. Leiva, and Anastasia Sergeeva. 2023. Are You Sure You Want to Order That? On Appropriateness of Voice-Only Proactive Feedback Strategies. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) (CUI '23). Association for Computing Machinery, New York, NY, USA, Article 43, 6 pages. <https://doi.org/10.1145/3571884.3604312>
- [21] W Jack Duncan. 1982. Humor in management: Prospects for administrative practice and research. *Academy of management review* 7, 1 (1982), 136–142.
- [22] Paweł Dybala, Michał Ptaszynski, Rafał Rzepka, and Kenji Araki. 2009. Humoroids: conversational agents that induce positive emotions with humor. In *AAMAS'09 Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, Vol. 2. ACM, ACM, New York, NY, USA, 1171–1172.
- [23] Justin Edwards, Christian Janssen, Sandy Gould, and Benjamin R. Cowan. 2021. Eliciting Spoken Interruptions to Inform Proactive Speech Agent Design. In *CUI 2021 - 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) (CUI '21). Association for Computing Machinery, New York, NY, USA, Article 23, 12 pages. <https://doi.org/10.1145/3469595.3469618>
- [24] Jonathan B Evans, Jerel E Slaughter, Aleksander PJ Ellis, and Jessi M Rivin. 2019. Gender and the evaluation of humor at work. *Journal of Applied Psychology* 104, 8 (2019), 1077.
- [25] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019. A taxonomy of social cues for conversational agents. *International Journal of Human-Computer Studies* 132 (2019), 138–161.
- [26] Kennan Ferguson. 2019. Comedy and critical thought: Laughter as resistance: Krista Bonello, Rutter Giappone, Fred Francis, and Iain MacKenzie (Eds.) London: Rowman and Littlefield International, 2018, x+ 237pp., ISBN: 978-1786604071.
- [27] Gabriella Gandino, M Vesco, S Ramella Benna, M Prastaro, et al. 2010. Whiplash for the Mind. Humour in Therapeutic Conversation. *International Journal of Psychotherapy* 14 (2010), 13–24.
- [28] Xiang Ge, Dan Li, Daisong Guan, Shihui Xu, Yanyan Sun, and Moli Zhou. 2019. Do smart speakers respond to their errors properly? A study on human-computer dialogue strategy. In *Design, User Experience, and Usability. User Experience in Advanced Technological Environments: 8th International Conference, DUXU 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part II 21*. Springer, Springer, Cham, 440–455.
- [29] Eun Go and S Shyam Sundar. 2019. Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior* 97 (2019), 304–316.
- [30] Talip Gonulal. 2021. Investigating EFL learners' humorous interactions with an intelligent personal assistant. *Interactive Learning Environments* 0, 0 (2021), 1–14. <https://doi.org/10.1080/10494820.2021.1974489>
- [31] Christian F Hempelmann. 2008. Computational humor: Beyond the pun? *The Primer of Humor Research. Humor Research* 8 (2008), 333–360.
- [32] Hsiu-Fang Hsieh and Sarah E Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative health research* 15, 9 (2005), 1277–1288.
- [33] Scott Hudson, James Fogarty, Christopher Atkeson, Daniel Avrahami, Jodi Forlizzi, Sara Kiesler, Johnny Lee, and Jie Yang. 2003. Predicting Human Interruptibility with Sensors: A Wizard of Oz Feasibility Study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) (CHI '03). Association for Computing Machinery, New York, NY, USA, 257–264. <https://doi.org/10.1145/642611.642657>
- [34] Kristiina Jokinen and Michael McTear. 2009. Spoken dialogue systems. *Synthesis Lectures on Human Language Technologies* 2, 1 (2009), 1–151.

- [35] A Kelly. 2017. Siri, tell me a joke. No, not that one. Could machine learning help the voice-activated assistant find its comedic chops. *Signal* 71, 7 (2017), 11–12.
- [36] Chloé Kiddon and Yuriy Brun. 2011. That's What She Said: Double Entendre Identification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2* (Portland, Oregon) (*HLT '11*). Association for Computational Linguistics, USA, 89–94.
- [37] Rafal Kocielnik, Daniel Avrahami, Jennifer Marlow, Di Lu, and Gary Hsieh. 2018. Designing for Workplace Reflection: A Chat and Voice-Based Conversational Agent. In *Proceedings of the 2018 Designing Interactive Systems Conference* (Hong Kong, China) (*DIS '18*). Association for Computing Machinery, New York, NY, USA, 881–894. <https://doi.org/10.1145/3196709.3196784>
- [38] Mitsuki Komori, Yuichiro Fujimoto, Jianfeng Xu, Kazuyuki Tasaka, Hiromasa Yanagihara, and Kinya Fujita. 2019. Experimental Study on Estimation of Opportune Moments for Proactive Voice Information Service Based on Activity Transition for People Living Alone. In *Human-Computer Interaction. Perspectives on Design*, Masaaki Kurosu (Ed.). Springer International Publishing, Cham, 527–539.
- [39] Dejun Tony Kong, Cecily D Cooper, and John J Sosik. 2019. The state of research on leader humor. *Organizational psychology review* 9, 1 (2019), 3–40.
- [40] Matthias Kraus, Marvin Schiller, Gregor Behnke, Pascal Bercher, Michael Dorna, Michael Dambier, Birte Glimm, Susanne Biundo, and Wolfgang Minker. 2020. "Was That Successful?" On Integrating Proactive Meta-Dialogue in a DIY-Assistant Using Multimodal Cues. In *Proceedings of the 2020 International Conference on Multimodal Interaction* (Virtual Event, Netherlands) (*ICMI '20*). Association for Computing Machinery, New York, NY, USA, 585–594. <https://doi.org/10.1145/3382507.3418818>
- [41] Tracy Kubert and Elena Korshakova. 2020. Identifying IPA humor styles. *Proceedings of the Association for Information Science and Technology* 57, 1 (2020), e411.
- [42] NICHOLAS A. KUIPER and CATHERINE LEITE. 2010. Personality impressions associated with four distinct humor styles. *Scandinavian Journal of Psychology* 51, 2 (2010), 115–122. <https://doi.org/10.1111/j.1467-9450.2009.00734.x> arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9450.2009.00734.x>
- [43] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, Are You Listening? Privacy Perceptions, Concerns and Privacy-Seeking Behaviors with Smart Speakers. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 102 (Nov. 2018), 31 pages. <https://doi.org/10.1145/3274371>
- [44] Uichin Lee, Kyungsik Han, Hyunsung Cho, Kyong Mee Chung, Hwajung Hong, Sung Ju Lee, Youngtae Noh, Sooyoung Park, and John M. Carroll. 2019. Intelligent positive computing with mobile, wearable, and IoT devices: Literature review and research directions. *Ad Hoc Networks* 83 (Feb. 2019), 8–24. <https://doi.org/10.1016/j.adhoc.2018.08.021> Funding Information: This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2017M3C4A7083529). Publisher Copyright: © 2018 Elsevier B.V..
- [45] Richard G Lomax and Seyed A Moosavi. 2002. Using humor to teach statistics: Must they be orthogonal? *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences* 1, 2 (2002), 113–130.
- [46] Ivan Lombardi. 2012. Not-so-serious games for language learning. Now with 99, 9% more humour on top. *Procedia Computer Science* 15 (2012), 148–158.
- [47] Irene Lopatovska, Pavel Braslavski, Alice Griffin, Katherine Curran, Armando Garcia, Mary Mann, Alexandra Srp, Sydney Stewart, Alanood Al Thani, Shannon Mish, Wanyi Wang, and Monica G. Maceli. 2020. Comparing Intelligent Personal Assistants on Humor Function. In *Sustainable Digital Communities*, Anneli Sundqvist, Gerd Berget, Jan Nolin, and Kjell Ivar Skjerdingsstad (Eds.). Springer International Publishing, Cham, 828–834.
- [48] Irene Lopatovska, Elena Korshakova, and Tracy Kubert. 2020. Assessing user reactions to intelligent personal assistants' humorous responses. *Proceedings of the Association for Information Science and Technology* 57, 1 (2020), e256.
- [49] Jackson G. Lu, Ashley E. Martin, Anastasia Usova, and Adam D. Galinsky. 2019. Chapter 9 - Creativity and Humor Across Cultures: Where Aha Meets Haha. In *Creativity and Humor*, Sarah R. Luria, John Baer, and James C. Kaufman (Eds.). Academic Press, Cambridge, Massachusetts, 183–203. <https://doi.org/10.1016/B978-0-12-813802-1.00009-0>
- [50] Jim Lyttle. 2001. The effectiveness of humor in persuasion: The case of business ethics training. *The Journal of general psychology* 128, 2 (2001), 206–216.
- [51] Nathan Malkin, Joe Deatrick, Allen Tong, Primal Wijesekera, Serge Egelman, and David Wagner. 2019. Privacy attitudes of smart speaker users. *Proceedings on Privacy Enhancing Technologies* 2019, 4 (2019), 250–271.
- [52] Rod A Martin, Nicholas A Kuiper, L Joan Olinger, and Kathryn A Dance. 1993. Humor, coping with stress, self-concept, and psychological well-being.
- [53] Rod A Martin, Patricia Puhlik-Doris, Gwen Larsen, Jeanette Gray, and Kelly Weir. 2003. Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality* 37, 1 (2003), 48–75. [https://doi.org/10.1016/S0092-6566\(02\)00534-2](https://doi.org/10.1016/S0092-6566(02)00534-2)
- [54] Peter McGraw, Philip Fernbach, and Julie Schiro. 2012. All Kidding Aside: Humor Lowers Propensity to Remedy a Problem.
- [55] O. Miksik, I. Munasinghe, J. Asensio-Cubero, S. Reddy Bethi, S-T. Huang, S. Zylfo, X. Liu, T. Nica, A. Mitrosak, S. Mezza, R. Beard, R. Shi, R. Ng, P. Mediano, Z. Fountas, S-H. Lee, J. Medvesek, H. Zhuang, Y. Rogers, and P. Swietojanski. 2020. Building Proactive Voice Assistants: When and How (not) to Interact. arXiv:2005.01322 [cs.HC]
- [56] Christopher Mims. 2016. Your Next Friend Could Be a Robot—WSJ [News].
- [57] R Narula, V Chaudhary, K Narula, and R Narayan. 2011. Depression, anxiety and stress reduction in medical education: Humor as an intervention. *Online J Health Allied Scs* 10, 1 (2011), 7.
- [58] Anton Nijholt, Andreea Niculescu, Alessandro Valitutti, and Rafael E. Banchs. 2017. Humor in Human-Computer Interaction: A Short Survey. In *Adjunct Proceedings INTERACT 2017 Mumbai*, Anirudha Joshi, Devanuj K. Balkrishan, Girish Dalvi, and Marco Winckler (Eds.). Indian Institute of Technology Madras, India, 192–214.
- [59] Joseph Polimeni and Jeffrey P Reiss. 2006. The first joke: Exploring the evolutionary origins of humor. *Evolutionary psychology* 4, 1 (2006), 147470490600400129.
- [60] Aung Pyae and Tapani N. Joellson. 2018. Investigating the Usability and User Experiences of Voice User Interface: A Case of Google Home Smart Speaker. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Barcelona, Spain) (*MobileHCI '18*). Association for Computing Machinery, New York, NY, USA, 127–131. <https://doi.org/10.1145/3236112.3236130>
- [61] Ashwin Rajadesingan, Reza Zafarani, and Huan Liu. 2015. Sarcasm Detection on Twitter: A Behavioral Modeling Approach. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (Shanghai, China) (*WSDM '15*). Association for Computing Machinery, New York, NY, USA, 97–106. <https://doi.org/10.1145/2684822.2685316>
- [62] Brandy Reece. 2014. Putting the Ha! In Aha!: Humor as a Tool for Effective Communication.
- [63] Leon Reicherts, Yvonne Rogers, Licia Capra, Ethan Wood, Tu Dinh Duong, and Neil Sebire. 2022. It's Good to Talk: A Comparison of Using Voice Versus Screen-Based Interactions for Agent-Assisted Tasks. *ACM Trans. Comput.-Hum. Interact.* 29, 3, Article 25 (jan 2022), 41 pages. <https://doi.org/10.1145/3484221>
- [64] Leon Reicherts, Nima Zargham, Michael Bonfert, Yvonne Rogers, and Rainer Malaka. 2021. May I Interrupt? Diverging Opinions On Proactive Smart Speakers. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) (*CUI '21*). Association for Computing Machinery, New York, NY, USA, Article 34, 10 pages. <https://doi.org/10.1145/3469595.3469629>
- [65] Graeme Ritchie. 2009. Can computers create humor? *AI Magazine* 30, 3 (2009), 71–71.
- [66] Hannes Ritschel, İlhan Aslan, David Sedlbauer, and Elisabeth André. 2019. Irony man: Augmenting a social robot with the ability to use irony in multimodal communication with humans.
- [67] A Joy Rivera. 2014. A socio-technical systems approach to studying interruptions: Understanding the interrupter's perspective. *Applied ergonomics* 45, 3 (2014), 747–756.
- [68] Eric J Romero and Kevin W Cruthirds. 2006. The use of humor in the workplace. *Academy of management perspectives* 20, 2 (2006), 58–69.
- [69] Maaik Roubroeks, Jaap Ham, and Cees Midden. 2011. When Artificial Social Agents Try to Persuade People: The Role of Social Agency on the Occurrence of Psychological Reactance. *International Journal of Social Robotics* 3, 2 (Apr 2011), 155–165. <https://doi.org/10.1007/s12369-010-0088-1>
- [70] Maaik Roubroeks, Cees Midden, and Jaap Ham. 2009. Does it make a difference who tells you what to do?: exploring the effect of social agency on psychological reactance. In *Proceedings of the 4th International Conference on Persuasive Technology*. ACM, Claremont California USA, 1–6. <https://doi.org/10.1145/1541948.1541970>
- [71] Maria Schmidt and Patricia Braunger. 2018. A Survey on Different Means of Personalized Dialog Output for an Adaptive Personal Assistant. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization* (Singapore, Singapore) (*UMAP '18*). Association for Computing Machinery, New York, NY, USA, 75–81. <https://doi.org/10.1145/3213586.3226198>
- [72] Maria Schmidt, Wolfgang Minker, and Steffen Werner. 2020. User Acceptance of Proactive Voice Assistant Behavior. In *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2020*, Andreas Wendemuth, Ronald Böck, and Ingo Siegert (Eds.). TUDpress, Dresden, Dresden, Germany, 18–25.
- [73] Maria Schmidt, Daniela Stier, Steffen Werner, and Wolfgang Minker. 2019. Exploration and assessment of proactive use cases for an in-car voice assistant. In *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2019*, Peter Birkholz and Simon Stone (Eds.). TUDpress, Dresden, Dresden, Germany, 148–155.
- [74] Andreea C Schöpf, Gillian S Martin, and Mary A Keating. 2017. Humor as a communication strategy in provider–patient communication in a chronic care setting. *Qualitative Health Research* 27, 3 (2017), 374–390.

- [75] Chen Shani, Alexander Libov, Sofia Tolmach, Liane Lewin-Eytan, Yoelle Maarek, and Dafna Shahaf. 2022. “Alexa, Do You Want to Build a Snowman?” Characterizing Playful Requests to Conversational Agents. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 423, 7 pages. <https://doi.org/10.1145/3491101.3519870>
- [76] Ben Sheehan, Hyun Seung Jin, and Udo Gottlieb. 2020. Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research* 115 (2020), 14–24.
- [77] Hyunju Shin, Isabella Bunosso, and Lindsay R. Levine. 2023. The influence of chatbot humour on consumer evaluations of services. *International Journal of Consumer Studies* 47, 2 (2023), 545–562. <https://doi.org/10.1111/ijcs.12849> arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/ijcs.12849>
- [78] Heung-Yeung Shum, Xiao-dong He, and Di Li. 2018. From Eliza to Xiaolce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering* 19 (2018), 10–26.
- [79] Timothy Sohn, Kevin A. Li, Gunny Lee, Ian Smith, James Scott, and William G. Griswold. 2005. Place-Its: A Study of Location-Based Reminders on Mobile Phones. In *Proceedings of the 7th International Conference on Ubiquitous Computing* (Tokyo, Japan) (UbiComp '05). Springer-Verlag, Berlin, Heidelberg, 232–250. https://doi.org/10.1007/11551201_14
- [80] Christina Steindl, Eva Jonas, Sandra Sittenthaler, Eva Traut-Mattausch, and Jeff Greenberg. 2015. Understanding Psychological Reactance: New Developments and Findings. *Zeitschrift für Psychologie* 223, 4 (Oct 2015), 205–214. <https://doi.org/10.1027/2151-2604/a000222>
- [81] Petra-Maria Strauss and Wolfgang Minker. 2010. *Proactive spoken dialogue interaction in multi-party environments*. Springer, Cham.
- [82] Madiha Tabassum, Tomasz Kosiński, Alisa Frik, Nathan Malkin, Primal Wijesekera, Serge Egelman, and Heather Richter Lipford. 2019. Investigating Users’ Preferences and Expectations for Always-Listening Voice Assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 4 (2019), 1–23.
- [83] Julia M Taylor and Victor Raskin. 2013. Towards the cognitive informatics of natural language: The case of computational humor. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)* 7, 3 (2013), 25–45.
- [84] Sascha Topolinski and Rolf Reber. 2010. Gaining insight into the “Aha” experience. *Current Directions in Psychological Science* 19, 6 (2010), 402–405.
- [85] Liam D. Turner, Stuart M. Allen, and Roger M. Whitaker. 2015. Interruption Prediction for Ubiquitous Systems: Conventions and New Directions from a Growing Field. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) (UbiComp '15). Association for Computing Machinery, New York, NY, USA, 801–812. <https://doi.org/10.1145/2750858.2807514>
- [86] Michel Vacher, Benjamin Lecouteux, Pedro Chahuara, François Portet, Brigitte Meillon, and Nicolas Bonnefond. 2014. The Sweet-Home speech and multimodal corpus for home automation interaction. , 4499–4506 pages.
- [87] Alessandro Valitutti, Hannu Toivonen, Antoine Doucet, and Jukka M Toivanen. 2013. “Let everything turn well in your wife”: generation of adult humor using lexical constraints. , 243–248 pages.
- [88] Alessandro Valitutti and Tony Veale. 2016. Infusing humor in unexpected events. In *Distributed, Ambient and Pervasive Interactions: 4th International Conference, DAPI 2016, Held as Part of HCI International 2016, Toronto, ON, Canada, July 17–22, 2016, Proceedings 4*. Springer, Springer, Cham, 370–379.
- [89] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. 2021. Eliciting and Analysing Users’ Envisioned Dialogues with Perfect Voice Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3411764.3445536>
- [90] Sarah Theres Völkel, Daniel Buschek, Malin Eiband, Benjamin R. Cowan, and Heinrich Hussmann. 2021. Eliciting and Analysing Users’ Envisioned Dialogues with Perfect Voice Assistants. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 254, 15 pages. <https://doi.org/10.1145/3411764.3445536>
- [91] Sarah Theres Völkel, Samantha Meindl, and Heinrich Hussmann. 2021. Manipulating and Evaluating Levels of Personality Perceptions of Voice Assistants through Enactment-Based Dialogue Design. In *Proceedings of the 3rd Conference on Conversational User Interfaces* (Bilbao (online), Spain) (CUI '21). Association for Computing Machinery, New York, NY, USA, Article 10, 12 pages. <https://doi.org/10.1145/3469595.3469605>
- [92] Christina Ziying Wei, Young-Ho Kim, and Anastasia Kuzminykh. 2023. The Bot on Speaking Terms: The Effects of Conversation Architecture on Perceptions of Conversational Agents. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) (CUI '23). Association for Computing Machinery, New York, NY, USA, Article 18, 16 pages. <https://doi.org/10.1145/3571884.3597139>
- [93] Jing Wei, Tilman Dingler, and Vassilis Kostakos. 2021. Developing the Proactive Speaker Prototype Based on Google Home. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–6.
- [94] Jing Wei, Tilman Dingler, and Vassilis Kostakos. 2022. Understanding User Perceptions of Proactive Smart Speakers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4, Article 185 (dec 2022), 28 pages. <https://doi.org/10.1145/3494965>
- [95] Marc G Weinberger and Charles S Gulas. 1992. The impact of humor in advertising: A review. *Journal of advertising* 21, 4 (1992), 35–59.
- [96] Alex C Williams, Harmanpreet Kaur, Gloria Mark, Anne Loomis Thompson, Shamsi T Iqbal, and Jaime Teevan. 2018. Supporting workplace detachment and reattachment with conversational intelligence. , 13 pages.
- [97] Kai Chi Yam, Christopher M Barnes, Keith Leavitt, Wu Wei, Jenson Lau, and Eric Luis Uhlmann. 2019. Why so serious? A laboratory and field investigation of the link between morality and humor. *Journal of Personality and Social Psychology* 117, 4 (2019), 758.
- [98] Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy. 2015. Humor recognition and humor anchor extraction. , 2367–2376 pages.
- [99] Xiaodong Yue, Feng Jiang, Su Lu, and Neelam Hiranandani. 2016. To be or not to be humorous? Cross cultural perspectives on humor. *Frontiers in psychology* 7 (2016), 1495.
- [100] Nima Zargham, Dmitry Alexandrovsky, Jan Erich, Nina Wenig, and Rainer Malaka. 2022. “I Want It That Way”: Exploring Users’ Customization and Personalization Preferences for Home Assistants. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (CHI '22). Association for Computing Machinery, New York, NY, USA, –.
- [101] Nima Zargham, Vino Avanesi, Leon Reicherts, Ava Elizabeth Scott, Yvonne Rogers, and Rainer Malaka. 2023. “Funny How?” A Serious Look at Humor in Conversational Agents. In *Proceedings of the 5th International Conference on Conversational User Interfaces* (Eindhoven, Netherlands) (CUI '23). Association for Computing Machinery, New York, NY, USA, Article 27, 7 pages. <https://doi.org/10.1145/3571884.3603761>
- [102] Nima Zargham, Mehrdad Bahrini, Georg Volkmar, Dirk Wenig, Karsten Sohr, and Rainer Malaka. 2019. What Could Go Wrong? Raising Mobile Privacy and Security Awareness Through a Decision-Making Game. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (Barcelona, Spain) (CHI PLAY '19 Extended Abstracts). Association for Computing Machinery, New York, NY, USA, 805–812. <https://doi.org/10.1145/3341215.3356273>
- [103] Nima Zargham, Leon Reicherts, Michael Bonfert, Sarah Theres Voelkel, Johannes Schoening, Rainer Malaka, and Yvonne Rogers. 2022. Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma. In *Proceedings of the 4th Conference on Conversational User Interfaces* (Glasgow, United Kingdom) (CUI '22). Association for Computing Machinery, New York, NY, USA, Article 3, 14 pages. <https://doi.org/10.1145/3543829.3543834>

A APPENDIX

Table 3: Median, IQR, and minimum and maximum values of Humor rating scores for each scenario.

Scenario	Mdn	IQR	Min	Max
Scenario 1: Meeting Reminder	2.50	4.00	1.00	7.00
Scenario 2: Health Risk	3.00	4.00	1.00	7.00
Scenario 3: Fact Checking	3.00	4.00	1.00	7.00
Scenario 4: Disagreement Clarification	3.00	2.50	1.00	7.00
Scenario 5: Nudging	3.00	3.25	1.00	7.00